



Advancing Real-Time Crime Weapon Detection and High-Risk Person Classification in Pre-Crime Scenes: A Comprehensive Machine Vision Approach Utilizing SSD Detector

By Taiwo. M. Akinmuyisitan, John Cosmas & Yusuff Adeniyi Giwa

Abstract- The application of state-of-the-art in deep learning detection algorithms, such as You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD), presents a significant opportunity for enhancing crime prevention and control strategies. This research focuses on leveraging the SSD algorithm to detect common crime weapons on individuals in both pre-crime video scenes and real-world crime scenarios. By thoroughly understanding the operational principles of the SSD algorithm, we adapted it for the identification of dangerous weapons commonly associated with violent crimes. Our detection model, which targets both weapons and individuals, establishes a robust foundation for an artificial intelligence (AI) system that accurately predicts individuals at high risk. The model first identifies the presence of a person and subsequently checks for any of the specified weapons. If a weapon is detected, the system further analyzes the individual's movement and speed within the frame of reference. Should the individual exceed a predetermined movement threshold, the system flags them as high risk.

Keywords: *single shot multipledetection (SDD), convolutional neural networks (cnn), deep learning, weapon detection. artificial intelligence, computer vision, image classification, machine learning, object detection, mean average precision.*

GJCST-D Classification: LCC Code: HV8079.C65



Strictly as per the compliance and regulations of:



Advancing Real-Time Crime Weapon Detection and High-Risk Person Classification in Pre-Crime Scenes: A Comprehensive Machine Vision Approach Utilizing SSD Detector

Taiwo. M. Akinmuyisitan ^a, John Cosmas ^σ & Yusuff Adeniyi Giwa ^p

Abstract- The application of state-of-the-art in deep learning detection algorithms, such as You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD), presents a significant opportunity for enhancing crime prevention and control strategies. This research focuses on leveraging the SSD algorithm to detect common crime weapons on individuals in both pre-crime video scenes and real-world crime scenarios. By thoroughly understanding the operational principles of the SSD algorithm, we adapted it for the identification of dangerous weapons commonly associated with violent crimes. Our detection model, which targets both weapons and individuals, establishes a robust foundation for an artificial intelligence (AI) system that accurately predicts individuals at high risk. The model first identifies the presence of a person and subsequently checks for any of the specified weapons. If a weapon is detected, the system further analyzes the individual's movement and speed within the frame of reference. Should the individual exceed a predetermined movement threshold, the system flags them as high risk. For this study, the SSD model utilized a VGG16 backbone and was trained on a dataset comprising 3,317 images, featuring four distinct weapon categories: handgun, shotgun, rifle, and knife. The dataset was collected from UCF, via Kaggle and complimented with additional weapons from Google download all, all the sources are secondary, open sources and loyalty-free. We achieved a mean average precision of 84.19% across five classes after training for 59 epochs. The findings of this research demonstrate the effectiveness of the SSD algorithm in crime prevention and control, contributing to the ongoing discourse surrounding the application of detection algorithms for crime prediction. This work aims to provide technological innovations that can assist local law enforcement agencies in their operational duties. Additionally, the insights gained from this study may enhance the detection of abnormal behavior within the broader field of artificial intelligence.

Keywords: *single shot multipledetection (SDD), convolutional neural networks (cnn), deep learning, weapon detection, artificial intelligence, computer vision, image classification, machine learning, object detection, mean average precision.*

I. INTRODUCTION

The alarming rise in crime rates globally, particularly in developing nations, poses significant challenges to public safety, peaceful assembly, and social

stability. These issues undermine the fundamental right to protection of life and the maintenance of order. However, advancements in technology, especially in emerging fields such as artificial intelligence, offer innovative solutions to address security gaps and bolster support for local law enforcement agencies [1]. The evolving nature of criminal activities has necessitated the development of sophisticated techniques to assist police in their duty to safeguard citizens and uphold social stability [2].

In recent years, the application of machine learning and computer vision for identifying crime hotspots, analyzing weapons of mass violence, and examining crime scenes has gained traction within academic circles, technology-driven nations, and various industries [3]. Among these techniques, the integration of computer vision and artificial intelligence for crime prediction has emerged as a promising approach. Single Shot Detection (SSD) has become one of the most effective model architectures for object detection, offering high accuracy and speed while adapting precision for real-time applications [4].

Single Shot Detection (SSD) is an object detection algorithm similar to You Only Look Once (YOLO), both of which utilize convolutional neural networks (CNN) in their architecture. It may be true to state that SSD is an advanced form of convolutional neural network (CNN) [5]. SSD facilitates efficient processing of images and videos (3D/4D tensors), making it particularly suitable for analyzing crime scenes and weapons, as well as being useful in manipulating the predicting of potential crime hotspots, including those that have yet to occur. Crime prediction is a sensitive endeavor that demands a high level of accuracy in representing observed situations; in many real-world scenarios, time is of the essence, and precise detection and representation are crucial [6].

This paper explores the adoption of single-shot detection (SSD) algorithms for crime prediction by leveraging extensive datasets of real-world crime footage, weapons, and reported incidents. The aim is to train an artificial intelligence (AI) model capable of predicting criminal activities through the analysis of visual data patterns. SSD has the potential to identify

Author: e-mail: taiwo.akinmuyisitan@brunel.ac.uk

potential criminal activities and alert law enforcement agencies, facilitating proactive measures. Its speed and accuracy make SSD particularly promising for real-time applications in this context

The study demonstrates the efficiency and effectiveness of single-shot detection (SSD) in various security-related domains, including license plate recognition, traffic monitoring, crime scene analysis, and real-time anomaly detection. However, its application specifically in crime hotspots and crime prediction, as well as its potential for identifying individuals at high public risk, remains largely unexplored, presenting significant opportunities for innovation and improvement. This paper aims to investigate the use of SSD for crime prediction, with a focus on its implementation and performance. Additionally, the research develops a Python scripts to leverage SSD as an advanced detection algorithm for classifying individuals deemed to be at significantly high risk.

The rest of the paper is organized as follows: Section II provides related work in crime prediction and object detection technologies. Section III describes the methodology used in this study, including data collection, preprocessing, and the SSD model architecture. Section IV presents the experimental results, analysis, and discussions of our findings. Finally, Section V presents our conclusion and recommendation including the research limitations and future work.

II. RELATED WORK

SSD remains an emerging research area for the prediction and detection of crime. This emerging research area is looking to build on the capacity of the single-shot multi-box algorithm, its performance, its success in object detection, and more noticeably its capabilities in striking a balance between speed, performance, and real-time application [4]. Single shot detection(SSD) is one of the most popular deep learning object detection frameworks that is known for its balance between speed and accuracy, it could be considered a fit for real-time applications for crime detection and prediction [1].

Recent research on SSD has been on achieving an increased speed while still improving its speed in detecting small objects; speed and accuracy are important in real-time applications and they are also very important parts of surveillance systems. For example, SSD's multi-scale feature maps and its default bounding boxes make it easy for it to handle different sizes and aspect ratios which are critical in detecting however small representation of scenes in criminal situations [9].

SSDs have shown great promise in crime prediction when used directly or indirectly with data sources such as close circuit TV (CCTV) and other sensory instruments for the identification of crime

hotspots, analysis of crime scenes, and prediction of crimes and criminal activities[10]. In a research carried out by [11], the author used SSD together with machine learning algorithms to analyze the patterns captured in surveillance footage and predict suspicious behaviors and people that present possible threats; this research demonstrates great promise in adopting object detection algorithms for crime prediction, especially in an urban environment where many events are happening simultaneously, synchronously and in real-time.

Moreover, SSD's architecture has been adapted to address the specific challenges of crime prediction, such as the need for high accuracy in detecting small objects (e.g., weapons) and the ability to process video streams in real time [4]. These adaptations include the use of advanced feature extraction techniques and the integration of SSD with other neural network models to improve its robustness and accuracy in challenging environments [12].

III. MATERIALS AND METHODS

This section outlines the methodologies used to achieve the results detailed in Section IV. The approaches are centered on the Single Shot MultiBox Detector (SSD) framework, which leverages a single deep neural network to predict bounding boxes and classify objects in images. Due to the specific nature of the study, our model was trained from near-scratch using a custom dataset designed to reflect real-world crime scenarios. This section further describes the hardware configurations, processes, and techniques involved in constructing custom object detectors specifically aimed at crime prediction.

a) Performance Metrics

Generally, measuring the performance of the object detector used mean average precision (mAP). The mean Average Precision (mAP) is a broadly used performance measures for evaluating the accuracy of object detection models in computer vision models. It combines three important aspects: precision, recall and F1, providing a comprehensive measure of a model's ability to locate and classify objects within an image. To compute mAP, precision-recall curves constructed for each class of objects in our dataset. These curves plot precision against recall at different confidence score thresholds. The area under the curves then averaged across all object classes, resulting in the mean average precision. The mean Average Precision is a valuable metric for assessing the accuracy of object detection models. It offers a balanced evaluation by considering precision and recall, making it particularly suitable for object detection and classification problems.

The mean average precision calculated with equation (1).

$$mAP = N^{-1} \sum_{i=1}^N AP_i \quad (1)$$

- N is the total number of object classes of the dataset
- AP_i is the average precision for each class i.

i. *Precision*

Precision measures the proportion of correctly identified positive detections out of all the detections made by the model. It assesses the model's accuracy in predicting true positives while minimizing false positives.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

TP= True Positive, which are the instances that are true objects and are positive by the model.

FP = False Positive, which are instances that are not true objects of a class but positive for the model

ii. *Recall*

It calculates the proportion of true positives detected by the model out of all the ground truth objects present in the image. It measures the model's ability to find all relevant objects, minimizing false negatives.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Here,

FN= False Negative. These are instances where the object is positive (criminal), but the model wrongly predicted (failed to identify the criminal).

iii. *System F1 Value*

The F1 score is the harmonic mean of precision and recall and provides a balance between the two metrics.

It is calculated as:

$$[F1 \text{ Score} = 2 \times \frac{\{Precision \times Recall\}}{\{Precision + Recall\}}] \\ \{2(precision \times Recall)\} / precision + Recall \quad \text{eqn (4)}$$

For simplicity,

- Precision focuses on the accuracy of positive predictions.
- Recall focuses on the proportion of actual positives that correctly identified.
- F1 score provides a balance between precision and recall, especially when there is an uneven class distribution.

b) *Data Collection and Processing*

i. *Data Acquisition*

Implementing an SSD-based crime prediction model necessitates a substantial and carefully curated dataset. For the training of this system, a secondary video dataset was used, simulating a live-action.

These scenarios were crafted to replicate real-world criminal behavior, enabling the model to be trained and tested on its ability to generalize across diverse threatening situations. The dataset was enhanced by sourcing images and videos of weapons from online platforms and integrating real-world crime data from the UCF-Crime dataset, which includes video clips labeled with tags such as "assault" and "robbery" [14]



Fig. 1: Shotgun Sample



Fig. 2: UCF-Dataset [14]

ii. *Data Annotation*

Accurate data labeling is crucial for training deep learning models with SSD. In this study, manual annotation was performed using *labellmg* which ensures precise bounding box placement and class labels. At every labeling instance i.e an object (Image), a corresponding XML file is automatically created which contains the object metadata as follows;

- Class Name
- Bounding box Coordinates (xmin, ymin, xmax, ymax)
- Object Dimensions (width, height) The design covered five classes of interest:
- Person
- Handgun
- Shotgun
- Knife
- Rifle

This structured data enabled SSD to detect and classify these objects during training [15].

While YOLOv4 has its annotation in a .txt file, Pascal annotation standard format is in .xml, Fig 3.0

```

<annotation>
  <folder>shgun</folder>
  <filename>person.jpeg</filename>
  <path>/home/shgun/person.jpeg</path>
  <source>
    <database>Unknown</database>
    </source>
  <size>
    <width>234</width>
    <height>215</height>
    <depth>3</depth>
    </size>
  <segmented>0</segmented>
  <object>
    <name>Rifle</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>18</xmin>
      <ymin>11</ymin>
      <xmax>103</xmax>
      <ymax>183</ymax>
    </bndbox>
  </object>
  <object>
    <name>Shotgun</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>171</xmin>
      <ymin>77</ymin>
      <xmax>197</xmax>
      <ymax>131</ymax>
    </bndbox>
  </object>
  <object>
    <name>Person</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>54</xmin>
      <ymin>59</ymin>
      <xmax>211</xmax>
      <ymax>205</ymax>
    </bndbox>
  </object>
</annotation>
    
```

Fig. 3: SDD Sample Dataset Image .XML Annotation file Format

In preparing a dataset for training, a folder named “SSD Custom” created then inside this folder, three folders created namely.

1. Annotations
2. Imageset
3. JPEGImages

c) System Framework

SSD, known for its efficiency in real-time object detection, processes entire images in a single forward pass through the network. This framework predicts multiple bounding boxes and associated class scores for each object, using feature maps extracted at different scales. The advantage of SSD lies in its speed and ability to detect objects of various sizes

simultaneously, making it ideal for applications where rapid response is essential, such as crime prediction [16].

The system architecture shown in Figure 4.0 utilizes SSD with vgg16 as the backbone network. The input data-drone-captured videos-are fed into the SSD, where the frames are divided into grids. Each grid cell then predicts bounding boxes and class scores, enabling the detection of objects relevant to criminal activities.

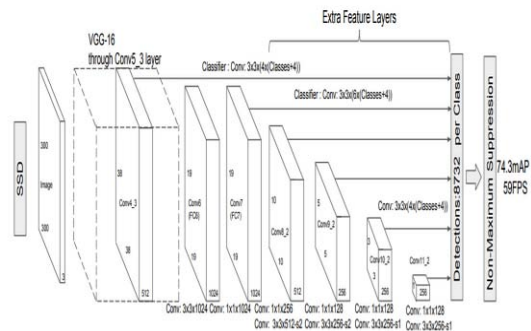


Fig. 4: SSD System Architecture [16]

d) Model Setup

We used SSD with vgg16 as the backbone. The Vgg16 was selected as the backbone model because of the depth of its architecture; Vgg16 has a depth of 16 convolutional layers which makes it easily learn complex and hierarchical features of our training dataset [19]. Additionally, vgg16 has done incredibly well in several image recognition tasks, it has achieved a very high accuracy over image classification tasks [17].

In this paper, we used mutliibox loss which has in its regression and classification loss. The regression loss measures the accuracy of the bounding box while classification predicts the performance of the model over the classes.

Training Configuration:

- **Batch Size:** There is a need to balance memory constraints and model convergence, therefore a small batch size of 10 was used .
- **Epochs:** The model is set for 100 epochs, with validation performed every epoch to monitor performance and avoid overfitting.
- **Checkpointing:** Models are saved at regular intervals, enabling the recovery of the best-performing model.
- **Optimization:** The model used stochastic gradient descent (SGD) which helps speed model convergence.

e) Evaluation

The primary evaluation metric is mean average precision which measures the model accuracy over all classes [18]. mAP is computed at different Intersections

over Union (IoU) thresholds to provide a comprehensive assessment of the model's detection capabilities.

IV. RESULTS AND DISCUSSION

The entirety of the research was jointly conducted on p2.xlarge AWS GPU and core i5 HP pavilion 16GB RAM with a 4GB Nvidia VRAM. The p2.xlarge instance was connected to a local machine. The research was implemented with a single shot multi-box detection, the VGG16 was selected as the backbone for this research due to its performance and accuracy on image classification tasks where object localizations are of utmost importance. The dataset was a mix of data from a UCF dataset of known criminals, and open-source data of images of guns, knives, and persons. The dataset was a mix of data from a UCF dataset of known criminals, and open-source data of images of guns, knives, and persons. The object dimension(Width, Height), coordinates and class they belong to were determined during the annotation of the dataset. The total dataset was 3317 and was divided into train, test, and evaluation in the 80:10:10 ratio. The dataset was loaded into the training in batches of 10 to ensure adequate memory usage and was trained over 62 epochs. The model performance (mean average precision, mAP) is shown in Fig. 5.0 while Fig 5.1 shows the training loss.

```

File Edit View Search Terminal Help
2024-08-14 08:08:55 - evaluating average precision Image 267 / 316
2024-08-14 08:08:55 - evaluating average precision Image 268 / 316
2024-08-14 08:08:55 - evaluating average precision Image 269 / 316
2024-08-14 08:08:55 - evaluating average precision Image 270 / 316
2024-08-14 08:08:55 - evaluating average precision Image 271 / 316
2024-08-14 08:08:55 - evaluating average precision Image 272 / 316
2024-08-14 08:08:56 - evaluating average precision Image 273 / 316
2024-08-14 08:08:56 - evaluating average precision Image 274 / 316
2024-08-14 08:08:56 - evaluating average precision Image 275 / 316
2024-08-14 08:08:56 - evaluating average precision Image 276 / 316
2024-08-14 08:08:56 - evaluating average precision Image 277 / 316
2024-08-14 08:08:56 - evaluating average precision Image 278 / 316
2024-08-14 08:08:56 - evaluating average precision Image 279 / 316
2024-08-14 08:08:57 - evaluating average precision Image 280 / 316
2024-08-14 08:08:57 - evaluating average precision Image 281 / 316
2024-08-14 08:08:57 - evaluating average precision Image 282 / 316
2024-08-14 08:08:57 - evaluating average precision Image 283 / 316
2024-08-14 08:08:57 - evaluating average precision Image 284 / 316
2024-08-14 08:08:57 - evaluating average precision Image 285 / 316
2024-08-14 08:08:57 - evaluating average precision Image 286 / 316
2024-08-14 08:08:58 - evaluating average precision Image 287 / 316
2024-08-14 08:08:58 - evaluating average precision Image 288 / 316
2024-08-14 08:08:58 - evaluating average precision Image 289 / 316
2024-08-14 08:08:58 - evaluating average precision Image 290 / 316
2024-08-14 08:08:58 - evaluating average precision Image 291 / 316
2024-08-14 08:08:58 - evaluating average precision Image 292 / 316
2024-08-14 08:08:59 - evaluating average precision Image 293 / 316
2024-08-14 08:08:59 - evaluating average precision Image 294 / 316
2024-08-14 08:08:59 - evaluating average precision Image 295 / 316
2024-08-14 08:08:59 - evaluating average precision Image 296 / 316
2024-08-14 08:08:59 - evaluating average precision Image 297 / 316
2024-08-14 08:08:59 - evaluating average precision Image 298 / 316
2024-08-14 08:08:59 - evaluating average precision Image 299 / 316
2024-08-14 08:09:00 - evaluating average precision Image 300 / 316
2024-08-14 08:09:00 - evaluating average precision Image 301 / 316
2024-08-14 08:09:00 - evaluating average precision Image 302 / 316
2024-08-14 08:09:00 - evaluating average precision Image 303 / 316
2024-08-14 08:09:00 - evaluating average precision Image 304 / 316
2024-08-14 08:09:00 - evaluating average precision Image 305 / 316
2024-08-14 08:09:01 - evaluating average precision Image 306 / 316
2024-08-14 08:09:01 - evaluating average precision Image 307 / 316
2024-08-14 08:09:01 - evaluating average precision Image 308 / 316
2024-08-14 08:09:01 - evaluating average precision Image 309 / 316
2024-08-14 08:09:01 - evaluating average precision Image 310 / 316
2024-08-14 08:09:01 - evaluating average precision Image 311 / 316
2024-08-14 08:09:02 - evaluating average precision Image 312 / 316
2024-08-14 08:09:02 - evaluating average precision Image 313 / 316
2024-08-14 08:09:02 - evaluating average precision Image 314 / 316
2024-08-14 08:09:02 - evaluating average precision Image 315 / 316
2024-08-14 08:09:04 - Average Precision Per-class:
2024-08-14 08:09:04 - Person: 0.88960919199411
2024-08-14 08:09:04 - Shotgun: 0.848538171453688
2024-08-14 08:09:04 - Handgun: 0.9158510911899285
2024-08-14 08:09:04 - Knife: 0.773551914978061
2024-08-14 08:09:04 - Rifle: 0.80281888702274093
2024-08-14 08:09:04 - Mean Average Precision (mAP): 0.8419425859090884
(torchSSD) shegun@shegun-HP-Pavillon-Gaming-Laptop-15-cx0xxx:~/jetson-tra
    
```

Fig. 5: Mean Average Precision (mAP)

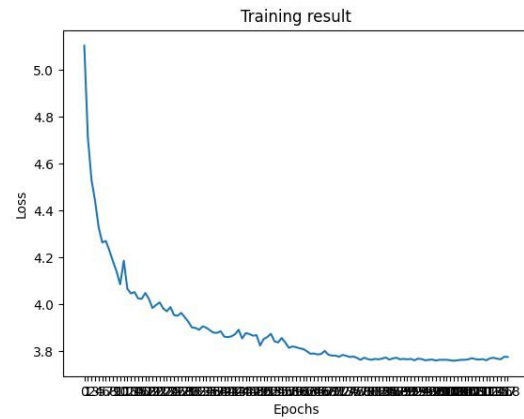


Fig. 5.1: Training Loss

The model was trained and initially set at 100 epochs. However, at about 62 epochs, the model began to overfit, an indication that the model isn't learning anything again. As such, it was ended. The best model performance therefore achieved at 59 epochs with a mean average precision of 0.84%. The inference was conducted on the data shown in Fig. 5(a) and the result of this is shown in Fig. 5(b).



Fig. 5(a): Sample Inference data



Fig. 5(b): Inference result

Table 1.0: SDD Class by Class Performance

Class	AP
Person	86.89
Shotgun	84.85
Handgun	91.58
Knife	77.35
Rifle	80.28

For our result, a confidence threshold of 0.25 was set, this threshold means an object to be detected, must first cross a 25% detection threshold for example for a class "Shotgun" to be said to be in a scene, the model must have seen the class crossed 25% initial detection threshold. This is to ensure that less confidence predictions are filtered out which therefore ensures a reliable detection mechanism.

In Fig 5.0, the model achieved a mean average precision of 84.19%. The high average precision (AP) for weapons mostly associated with violent crimes is of great significance to this research. To detect a person of high risk, it was important to have a model that would be able to pick features from scenes that may suggest whether or not a person detected is a person of high risk. The high precision results recorded in this phase (detection) were very instrumental in building a model that detects persons of high risk.

The High-Risk Detection System functions by identifying and flagging individuals who display potentially dangerous behavior or possess hazardous objects, utilizing a combination of speed analysis and object detection techniques. The process begins with the detection and tracking of "Person" instances within a video stream, where bounding boxes are drawn around each identified individual. The system uses Deep-SORT multiple object tracking algorithm. It calculates the speed of movement for each detected person by comparing their positions across consecutive frames. If an individual's speed exceeds a predefined threshold (set at 0.35 in the design Python script), they are flagged as "High Risk." Furthermore, the system also assesses the presence of specific high-risk objects, such as firearms (e.g., rifles, shotguns, handguns) and knives, associated with the detected "Person" in the precrime video footage.



Fig. 6(a): CCTV Crime Footage



Fig. 6(b): High Risk Prediction

The SSD (Single Shot MultiBox Detector) results demonstrate the efficacy of this approach. With an average precision (AP) of 86.89% for detecting "Person" instances, the model reliably identifies individuals, which is crucial for accurately determining whether someone should be flagged as high risk based on their behavior or the objects they carry. The model also shows strong performance in detecting specific weapons, with APs of 84.85% for "Shotgun" and 91.58% for "Handgun," which enhances the system's ability to identify individuals carrying these dangerous items. Fig. 6(a) and 6(b) show respectively the sample footage used for inference and the result. The footage is a real-life robbing scene captured by CCTV.

This model achieved high average precision (AP) scores for objects like "Handgun" (91.58%) and "Shotgun" (84.85%), However, other studies suggested that models like YOLOv4, which were fine-tuned for specific tasks and utilized additional enhancements, outperform our SSD in precision metrics, particularly for smaller objects [19]

V. CONCLUSION

This research focused on the application of the Single Shot Multi-Box Detection (SSD) algorithm for the prediction and prevention of violent crimes. It utilized a local HP Pavilion gaming machine with specific hardware specifications, alongside a p2.xlarge GPU instance on AWS, and a dataset comprising 3,317 instances from various sources, including the UCF Crime Open Dataset. We successfully developed a system capable of identifying individuals at high risk based on the presence of weapons and the speed at which these individuals are moving, indicating potential abnormal behavior.

To achieve the results presented in this paper, the datasets were meticulously annotated and divided into three distinct subsets: training, validation, and testing. The model was initially set to run for 100 epochs but was terminated at 62 epochs to prevent overfitting. Peak performance was achieved at 59 epochs, with a mean average precision of 84.19%.

In this study, we successfully employed the SSD algorithm to create a system for predicting high-risk individuals. This high-risk model relied heavily on the SSD model's performance in detecting weapons associated with violent crime. Upon optimization, our detection model demonstrated competitive results compared to the Yolov4 detection outcomes reported by [20], using the same dataset and classes.

The significance of our findings extends beyond the immediate results. Our model's ability to predict high-risk scenarios is particularly valuable in real-world surveillance applications, providing a robust tool for enhancing security measures. By successfully integrating advanced deep learning techniques, our study contributes to the literature on anomaly detection, especially concerning crime-related behaviors captured in surveillance videos.

However, our study is not without limitations. The dataset, while diverse, may still not cover all possible real-world scenarios, affecting the model's generalizability. Additionally, computational constraints limited the scale of our experiments. Future research could focus on expanding the dataset, improving annotation techniques, and exploring more powerful computational resources to further enhance model performance.

REFERENCES RÉFÉRENCES REFERENCIAS

1. W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 21-37.
2. Y. Chen, Y. Zhu, C. Ke, and F. Chen, "Crime prediction using deep learning," *IEEE Access*, vol. 7, pp. 150503-150510, 2019.
3. S. Zhang, R. Benenson, and B. Schiele, "CityPersons: A Diverse Dataset for Pedestrian Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4457-4465.
4. "What Is SSD & How It Improved Computer Vision Forever," *Analytics India Magazine*, 2024. [Online]. Available: <https://analyticsindiamag.com>. Accessed: Aug. 12, 2024.
5. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," *ar5iv.org*, 2024. [Online]. Available: <https://ar5iv.org/abs/1907.05047>. Accessed: Aug. 12, 2024.
6. G. C. Oatley and B. W. Ewart, "Crimes analysis software: 'pins in maps', clustering and Bayes net prediction," *Expert Systems with Applications*, vol. 25, no. 4, pp. 569-588, 2003.
7. O. Z. Apene, N. V. Blamah, and G. I. O. Aimufua, "Advancements in Crime Prevention and Detection: From Traditional Approaches to Artificial Intelligence Solutions," *European Journal of Applied Science, Engineering and Technology*, vol. 2, no. 2, pp. 285-297, 2024.
8. L. Aziz, M. S. B. H. Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review," *IEEE Access*, vol. 8, pp. 170461-170495, 2020.
9. L. Cheng, Y. Ji, C. Li, X. Liu, and G. Fang, "Improved SSD network for fast concealed object detection and recognition in passive terahertz security images," *Scientific Reports*, vol. 12, no. 1, p. 12082, 2022.
10. R. Shenoy, D. Yadav, H. Lakhotiya, and J. Sisodia, "An intelligent framework for crime prediction using behavioural tracking and motion analysis," in *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2022, pp. 1-6.
11. R. M. Saji and N. V. Sobhana, "Real Time Object Detection Using SSD For Bank Security," in *IOP Conference Series: Materials Science and Engineering*, vol. 1070, no. 1, p. 012060, Feb. 2021, IOP Publishing.
12. Y. J. Hwang, J. G. Lee, U. C. Moon, and H. H. Park, "SSD-TSEFFM: New SSD using trident feature and squeeze and extraction feature fusion," *Sensors*, vol. 20, no. 13, p. 3630, 2020.
13. A. Juneja, S. Juneja, A. Soneja, and S. Jain, "Real time object detection using CNN based single shot detector model," *Journal of Information Technology Management*, vol. 13, no. 1, pp. 62-80, 2021.
14. W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6479-6488.
15. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Cham, Switzerland: Springer, 2014, pp. 740-755.
16. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I*, Cham, Switzerland: Springer International Publishing, 2016, pp. 21-37.
17. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
18. R. Padilla, S. L. Netto, and E. A. Da Silva, "A Survey on Performance Metrics for ObjectDetection Algorithms," in *Proceedings of the 2020 International*

Conference on Systems, Signals and Image Processing (IWSSIP), 2020, pp. 237-242.

19. K. Tran and N. D. Tran Pham, "Comparative Analysis of Image Processing Object Detection Models: SSD MobileNet and YOLO for Guava Application," in *The International Conference on Sustainable Energy Technologies*, Singapore, Nov. 2023, pp. 443-451. Springer Nature Singapore.
20. Taiwo Akinmuyisitan and John Cosmas(2024) Enhanced Crime Prediction with Computer Vision: YOLOv4 Approach. *Global Journal of Computer Science and Technology*, Massachusset, USA Vol 24 issue(1).

